

Analysis of the social effort in multiplex participatory networks

Davide Vega, Roc Meseguer and Felix Freitag

Computer Architecture Department
Universitat Politecnica de Catalunya - BarcelonaTech
`{dvega,meseguer,felix}@ac.upc.edu`

Abstract. Community networks are participatory connectivity solutions for citizens where all the resources are owned, managed and controlled by the participants. As a natural evolution, in recent years some initiatives have flourished to provide higher level services based on volunteer computing and resource sharing paradigms. A fundamental aspect of these paradigms is user participation. In this work, we apply some social mining techniques aiming to identify the roles of the individuals in the social network behind a community network, here Guifi.net, and to measure the participatory involvement in the community network from 2003 to 2014. We observed that community network participants generally dedicate their time and effort to a single participatory forum, generating several types of community structures. We analyzed such structures using a multiplex network formed by mailing list in Guifi.net and a relationship graph built pairwise of users that share a physical wireless link. We were able to distinguish between non-hierarchical participatory forums, where almost all users are part of the same big community and two-tier participatory forums led by a small number of users that act as social bridges between their members. Finally, by testing the impact of community leaders in all participatory layers, we profiled the utility of the members' effort to the whole wireless community network.

Keywords: social effort, multiplex, collectives

1 Introduction

Community networks are growing fast as a sustainable model for self-provisioned computer networking infrastructures [6], alternative to other service offerings. This has been accelerated by the reduction of the costs of WiFi and optic networking equipment, combined with the growing popularity of wireless devices, and the lower complexity of network setup. In recent years, a plethora of non-profit initiatives have flourished to create community networks providing, among other services, Internet access. A few examples are Guifi.net [2] and FunkFeuer [1].

A characteristic of these initiatives is that the network topology grows organically, without a planned deployment or any consideration other than connecting

devices from new participants or locations by linking them to an existing one or improving the network service. Typically the deployment and management tasks are performed by the community network members, mostly volunteers. Avonts et.al. [3] reported that community networks members considered finding and keeping volunteers the largest organizational challenge, next to funding and finding devices maintainers the most second important challenge.

Beyond Internet access provision, the community networks' physical infrastructure is sometimes used by some members to provide applications (e.g. web servers, monitoring systems). As a natural evolution, some community networks are looking for ways to implement higher level applications [13], which would require mechanisms to regulate and normalize how their members interact with the computational resources [16]. The feasibility of implementing such contributory systems is highly dependent on the network participants' ability to rank and evaluate members' participation.

User participation can be measured in multiple ways and it is not limited to deploy and maintain physical devices or links. Community networks usually have other participatory forums where users can contribute to the growth and improvement of the network. Some community networks maintain online discussion forums while others use mailing lists or organize face-to-face meeting activities. These forums help users to organize and give support to new members to integrate into the network.

Fig. 1 shows the empirical cumulative distribution function (ECDF) plotted as the Lorenz curve, of users participation in Guifi.net, the largest community network to the best of our knowledge. The participation is measured separately as the number of new devices created by users and the number of messages exchanged in one of its participatory mailing list. The Gini coefficient [8], measured as the area between the line of equality and each of the curves, is close to the absolute inequality in both participatory forums – 0.8358 in the devices creation and 0.8320 in the message exchange. The Lorenz distribution function also suggests that network members behave differently in terms of participation in the examined forums.

As an example of individual different participatory involvement, we consider the number of messages and devices created by each user identified in both participatory forums in Fig. 2. We observe that most of the users are selective and choose to collaborate only in one of the participatory forums, contributing with little or nothing to the other. For example, there is a high concentration of users participating in the development mailing list, but these are users which contributed only with one or two devices to the physical communication network.

In this work we measure the level of user involvement in a wireless community network by applying social mining techniques to understand the differences observed among users of several participatory forums. We gather and process information from 13,407 registered users in the community network with more than 36,629 active communication devices, and a total of 10,045 threads and 49,355 messages in their most active mailing lists. We then study the evolution of participation and observe that participatory forums are currently in a mature

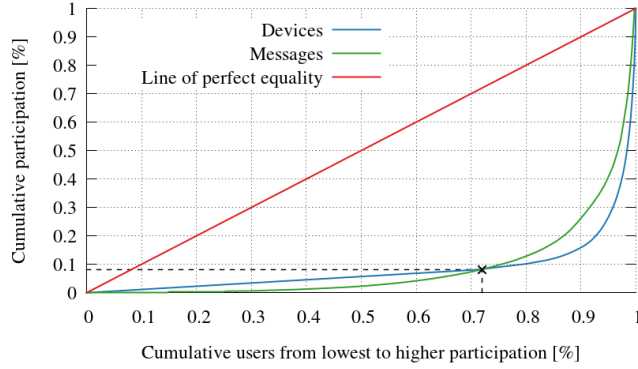


Fig. 1. Gini coefficient of two participatory forums in Guifi.net.

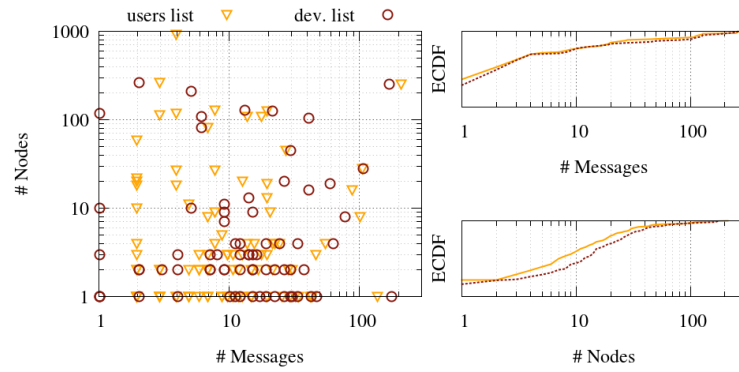


Fig. 2. Guifi.net users participation measured as the number of messages posted in the mailing lists users-list and dev-list, and the number of new communication devices.

state, which according to [10] suggests a relationship and subgroups analysis. Therefore, we conduct the analysis of interactions and community structure in each participatory forum separately. We find evidences of hierarchical community structures in the network's mailing lists with strong ties among their members and poor communication with other communities. We also observe a lack of structure in the communication network. Finally, we measure the significance of the members in each participatory layer to the community network as a global entity. Our methodology is inspired by [9], which conducted work on a multiplex and multi-layer analysis of online and offline users' interaction, where the authors discuss the existence of weak and strong ties among community members.

The contributions made in this work can be summarized as follows:

- We describe the evolution of members' activities since the creation of the network until 2014 (Section 3). We observed that the network is currently

in a mature stage and therefore the community and relationship analysis applies. Additionally we found evidence of seasonal participative patterns.

- We identify different community structures on each participatory layer, and observed that weak relations arise more in the communication layer, while strong relations are more common among members of the small communities that govern the mailing lists (Section 4.1).
- We measure the participants’ social value as individuals using centrality measures, and observed that only a small portion of the participants can be considered of having a high social value (Section 4.2).
- In terms of their impact on other layers, we simulate the robustness of the network against the disappearance of important members (Section 4.3), and found that their activities in each participatory forum should be measured differently, since their impact on the community network is also different.

2 Experimental framework

In this section we present our framework for the analysis of the social interactions and the estimation of the users’ effort in Guifi.net. We first introduce the Guifi.net wireless community network. Then, we describe the current participatory forums and how we aggregate their data to build our analysis graphs. Finally, we explain the assembly process used to build the multiplex graph.

2.1 Guifi.net network and information gathering

The Guifi.net network started in 2004 and in 2014 it has reached more than 24,000 operational devices, most of them in Catalonia and nearly all of them in the Iberian peninsula. In 2008 their members created the Guifi.net Foundation, a non-profit organisation responsible to coordinate the volunteers and provide deployment support to its users. The Guifi.net foundation encourages the deployment of the network, but does not control it. Therefore, virtually all the decisions concerning the network growth and maintenance are up to the users.

Physical nodes and communication layer representation: The network consists of a set of nodes interconnected through mostly wireless equipment that users must install and maintain, typically on building rooftops. The network grows driven by the needs of individuals. New links only appears based on the needs for connectivity by their owners or indirect beneficiaries (e.g. users of a community network crowd-funded by municipalities). Deploying a new node, or improving the connectivity for an existing one, is difficult without the cooperation of the owner/manager of the communication device to connect with. The communication layer intends to capture such interaction between participants in a graph structure to ease the analysis.

Using the topological graph built previously [16] and a dump of the Guifi.net web database, we relate each owner with their active devices as we show in Fig. 3. Then, we build an undirected graph where vertices represent the members of the network and edges represent a link in the topology graph that connects two nodes

Table 1. Summary of the basic properties of the graph layers

Network layer	Type	Nodes	Edges	Degree (max,min,avg.)	Components	Average distance	Diameter	Clustering coefficient
Communication	undirected	1919	3201	384,0,3.3361	250	3.238	7	0.1520
Users	directed	538	4607	521,0,17.1264	19	1.9675	2	0.751
Development	directed	401	3926	202,0,19.5810	13	2.3247	4	0.5614

owned or modified by one of the users. Additionally, the weight of the vertices represents the number of nodes created, while the weight of the edges stands for the number of links between users.

Mailing lists and social layer representation: Social participation in Guifi.net changed during the community network lifetime. Nowadays, mailing lists are the only online and social participatory forums left. The Guifi.net Foundation maintains two general-purpose mailing lists to coordinate users and developers since 2006. The first of them, *users-list*, is mainly used to discuss general topics, issues on coordinating physical infrastructure creation and maintenance and to help new users. The other one, *dev-list*, serves as a communication channel between some of the most active members in Guifi.net, most of them developers. Each mailing list is currently managed independently and contains only a small subset of the users registered in the web page – i.e. the dev-list reports 401 different users registered, while the Guifi.net webpage reports 13,407.

We used web scraping to gather information from the mailings lists dump published on the Guifi.net web site. Following the same methodology proposed by [7], for each mailing list we built a tree for each thread (see Fig. 3). Then, we used the sender email information to build a participants directed graph that includes as vertices the users in the mailing lists. While in the original framework the authors considered that connections between participants are reciprocal, we instead considered them directional. Therefore, in our graph an arc from user v to u , represents an answer from v to a message previously sent by u in the same thread, while arc’s weight stands for the amount of messages ever sent by v to u . We use vertices weight to keep track of the threads created by each user.

2.2 Homonymous detection

Homonymy is a characteristic of most distributed systems, like Peer-to-Peer applications, which implies the existence of users with multiple identifiers in the network [15]. Guifi.net anonymous exists because authentication is based on the email address of the participants and because the authentication for each participatory forum is independent. Homonymy was detected using the email similarity rule suggested by Bird et. al. [4], which is based on the Levenshtein edit distance between email address bases. Being as conservative as possible, we tag two identities as homonymous only if they have a distance of 1 or below.

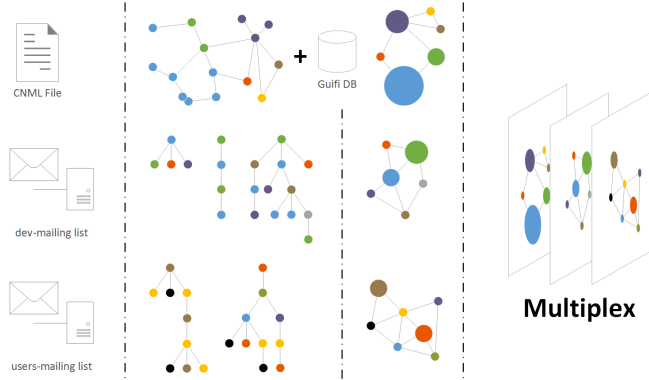


Fig. 3. The initial set of data gathered and the multiplex graph generation process.

2.3 Communities detection

Social relations in a group of participants can form a community if they are more willing to interact among them than with other members of the network. This is a well-known phenomena – called communities structure – that arises in most complex networks. The community’s size, structure or even members’ interactions outside and inside the communities are a good source of information to study the roles of the network users.

The community detection problem has been studied for a long time, and there are different algorithms and methods that can be applied, depending on the properties of the network and the properties of the targeted communities. In this work we apply two different methods, the *clique percolation method* [14] and the *Louvain method* [5] to detect two different community structures, and discuss the differences and the role of their members. In practical terms, the difference is that while the percolation method is based on the detection and aggregation of k -clique disjoint sets inside the graph – which will have maximum connectivity among their members –, the Louvain method is an optimized algorithm to find partitions providing that the modularity (the relationship between average degrees inside the community and intra communities) is minimized.

Multiplex graph: multiplex or multi-level graphs are abstract data structures which assemble the information of several graphs in such a way that each original graph is represented by a separated layer, meaning that each layer holds their original connectivity matrix [11]. The basic properties of each layer are summarized in Table 1. The structure of the multiplex enables to relate nodes with the same identifier between them. We built a 3-layer multiplex graph as the assemble of the participation layers (see Fig. 3). We use it to discuss the impact of the most active members of each layer in the whole network.

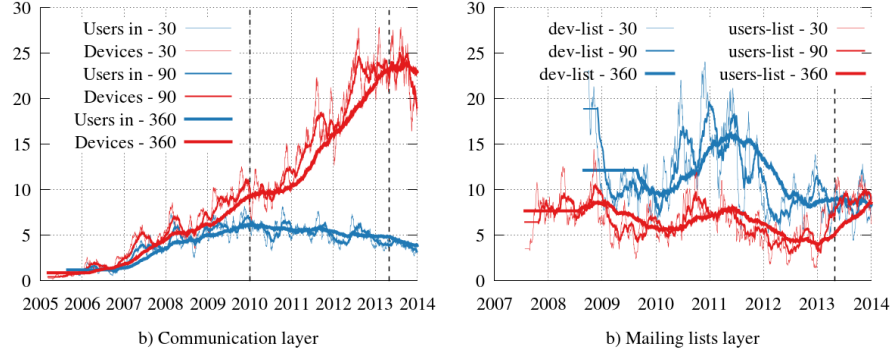


Fig. 4. Evolution of members' participation in each participatory layer.

3 Network evolution

In this section we study the activity performed by individuals on each participatory forum since the creation of the network. We analyze the activity of users by measuring the number of new physical communication devices registered in the network and the number of messages and threads sent in both mailing lists. Group activities are compared with the individual's interest, which lead us to the conclusion that Guifi.net is in a mature state with evidence of lack of interest by non-members.

3.1 Network state detection

Fig. 4 shows the activities done day by day by network participants in each participatory layer as the simple moving average at intervals of approximately one month (30), quarter of a year (90) and one year (360). We observe that users' participation switches from long periods of high activity to short periods of less activity. We identify the periods of lower activity, as those that correspond to the last quarter of every year – Spanish winter season.

The last cycle of less activity has been extended in the communication layer after a long period of increasing activity which lasted for 77 months. This change of trend was predictable as the number of new registered users have been decreasing the past 4 years. Nowadays only 20% of the new working devices are installed and managed by users registered 62 days before or less.

Regarding the activity in both mailing lists, it is not comparable because the number of members is quite different (25.46% difference according Table 1). However we can observe that, on average, the participation by user in the development list used to outnumber the participation in the users list (see Fig. 4).

3.2 Interest generated

Users' habits have changed over the time. While the interest in the communication network decreases since 2010, the users interest in the discussions and

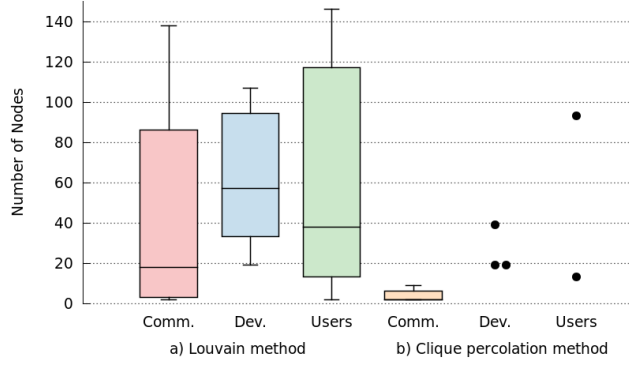


Fig. 5. Number of nodes by community. Comm., Dev. and Users refer to the communication, dev-list and users-list layers.

information exchange prevailed one year and a half after. Nowadays, while developers activities are decreasing, the rest of the users contribute to the increase of the users mailing list (see Fig. 4).

We conclude that voluntarism – measured as participation in the three layers – is stable in the network, and nowadays is concentrated around the senior members of the network. The network is still in a mature state, but is attracting fewer new users every day.

4 Communities and multiplex analysis

From the examination of the structure of the participatory forums in Guifi.net, we can consider the network mature enough to be subject of a community and interaction analysis. First, we focus on detecting possible independent communities which will capture the preferential interactions among members. Then, we address the detection of key actors in such communities to understand if there is some hierarchy. Finally, we address the multiplex analysis by understanding the role of layer authorities and well-connected members in the network.

4.1 Community structure

Community structure is a common characteristic shown by most complex networks, which allows us to discuss common properties among their members. We analyse the existence or not of community structures in the most important participatory forums, as a function of members interactions. Each boxplot in the Fig. 5 summarises the nodes composition of communities detected in our layers when the 2 different detection techniques are applied (see Section 2.3). Members of a layer which do not belong to any community are not represented.

We find that it is possible to divide all layers into several disjoint communities, where 75% of them have between 2 and 147 users. The communities median

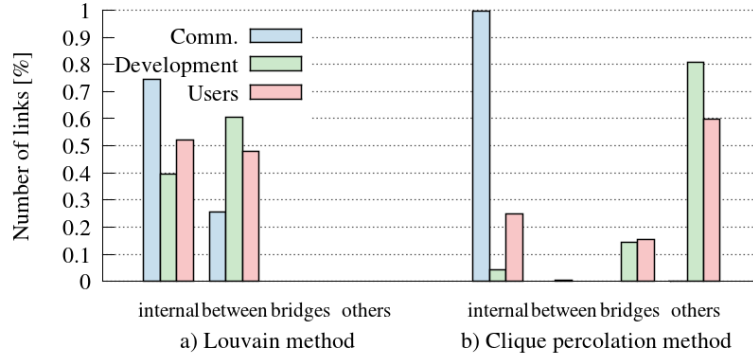


Fig. 6. Links distribution between communities. Bridges are links between two participants, both members of the same community, but one of them member of another community, too. Others are links to nodes which do not belong to any community.

size detected in the communication layer represents less than 1% of the users, while in the case of the development and users mailing lists it represents 14.4% and 7.11%, respectively, of the users.

In the communication layer 74.38% of the links are internal – between two members of the same community – and only 25.62% of the interactions are between members of two different communities, as shows the Fig. 6, resulting in a two-tier structure with users geographically close showing high participation among them. Participants in the mailing lists, however, show no preference in answering to members of their community or members of other communities.

The analysis of the participatory layers using the *clique percolation method* (see Section 2.3) reveals another community structure, enclosed by a core group of members. While the core group in the mailing list is formed by a small portion of users – from 17.28% in the users layer to 9.97% in the development layer, the core group of members 86.24% of the participants in the communication layer with 86.24% of the participants represents almost the entire network.

In contrast, the social layers show less ties among members of their communities than with other participants in the layer, as shown in Fig. 6, where communities are detected using cliques. Communities are connected only through bridges, suggesting that both social layers face a two-tier structure with several cores coordinated by some of the members. Finally, it is interesting to observe a stronger connectivity in the users mailing lists’ communities, which suggests a large gap of involvement between core community members and the other members of the participatory forum.

4.2 Individual participation and social value

We have seen so far that in the participatory forums of Guifi.net, 72% of the members generate more than 91.2% of the contributions. Nevertheless, the so-

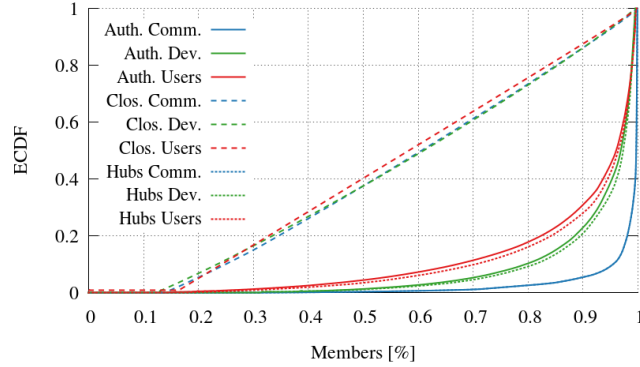


Fig. 7. Empirical Cumulative Distribution Function (ECDF) of members' centrality scores.

cial value of a community member for the whole network depends more on its connectivity and position towards the rest of the participants than on his or her individual contributions. For instance, if somebody sends a lot of messages to a single person, it does not imply that he or she is generating any social value.

We measured the individual impact of the users inside the network as their *closeness centrality*, *HITS authorities* and *hubs*. Closeness centrality for a connected graph is defined as the inverse of the average distance to all other nodes. In our case, the distance is measured as the sum of the weights of the nodes managed for the communication layer and of the messages answered in the mailing lists layers of each link it has to traverse.

HITS [12] is a ranking algorithm used in the past to exploit the web's hyper-link structure. As a result, we obtain two measures for each node, the authority and hubs. The first one is a measure of individuals as a source of information, while the second one ranks higher those nodes with high quality out-links. In order to apply the HITS algorithm directly to our mailing lists graphs, we transform them to undirected graphs.

Fig. 7 shows as we expected the same distribution for both HITS values in the communication layer due to the fact that the graph is undirected. Indeed, the social graphs present small differences between the authorities and hubs distributions. It is important to note that the authorities distribution shows that only a small portion of the users are of high value because their messages generate more replies from others.

Individuals' analysis revealed that there are only two users in common on all layers among the 10 higher ranked members, which are identified as the network founders. There are also two members in common in the mailing list which do not appear as top ranked in the communication layer. It highlights the higher affinity between members of the social participation layers compared with the communication layer.

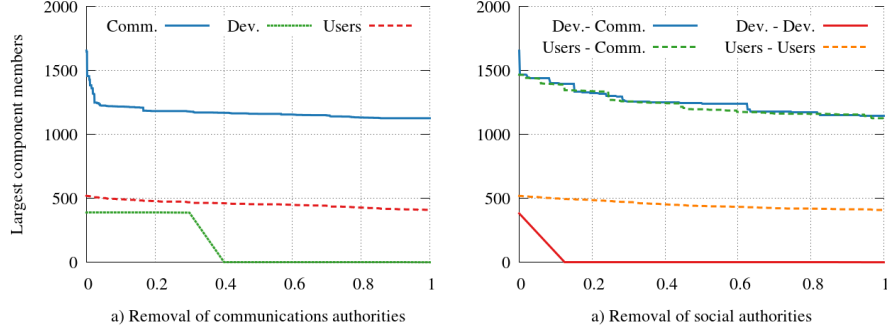


Fig. 8. Multiplex layers robustness.

4.3 Global impact of participation

In the previous section we discussed the social value generated by members in their own participatory layer. In this section, we turn our interest to understand their impact on the whole network by studying what happens when top layer authorities decide to leave the network. To this end, we firstly ordered the members of each participatory layer by the authorities score. Then, we proceeded recursively removing the top ranked node of the remaining ones and measuring the size of the largest component on each other layer in the multiplex.

Fig. 8 shows the robustness of each participatory layer when the authority members are recursively removed in the multiplex. Ideally, when an authority is removed, we expect that the size of each layer to be reduced only by one member – the authority itself. However, if the authority removed is essential to maintain the connectivity, we expect some other members getting disconnected – and the size of the layer’s biggest component reduced.

In examining the multiplex robustness in case of the disappearance of members from any layer, we observe a huge impact on the dev-list layer, and a moderate impact on the communication layer. However, the graph structure of the users-list layer makes it indifferent to these removals, even of their authorities. Thus, we can argue that the communication and dev-list layers show a higher dependence on the particular contributors.

5 Conclusions and discussion

5.1 Summary of contributions

This work makes several contributions to the analysis of member participation in a community network. Using the historical data from users’ activities in the participatory forums, we explained why the communities’ structure and multiplex analysis are relevant to measure members’ participation, by showing that our target network is in a mature state. Therefore, our findings could be applicable to other community networks in the same life-cycle stage.

We show that the analysed community structures indicate the existence of a hierarchical structure in the governance of the on-line participatory layers. This structure is led by a small set of members who only have contact with the rest of the subscribers through a few bridges. These bridge members are ranked highly as the most valuable members of the participatory layers. The asymmetry in participation found in the community network makes these members unique and essential for the communication network survival. Furthermore, the communication layer is structured in a weak way as an effect of the demographic distribution of their members.

5.2 Limitation of the analysis

Studying participation in the context of building new devices is relevant for measuring the effort dedicated to increase and improve the communication network, and as we have observed in this work, not all users dedicate the same effort to this task. However, there exist other ways to improve or to help building the physical network that were not captured by our model, because they require a very deep understanding of users' choices, and which could be more community network specific.

In general, the analysis of the social participatory forums allows detecting communities and users' roles. In our work, we conducted the study using the two most important mailing lists in Guifi.net network. However, many other social interactions occur through other ways, like the general face-to-face assembly of the Guifi.net foundation, sporadic physical encounters among members, or region specific mailing lists. Including the data from other mailing lists would lead us to understand in more detail the heterogeneity of the network and to detect more relevant communities in such regions. Nevertheless, with our current analysis we were able to detect clusters of users, whose actions and connections have a high impact on the whole community network.

5.3 Implications of the analysis

Most previous works have addressed the resource sharing or service regulation problem in various scenarios and measure the users' participation with physical resources. In a cooperative complex ecosystem like community networks, the effort dedicated to maintain and improve the network must be taken into account, too.

The social value analysis in Section 4.2 is a first step to measure the effort dedicated by community network members through their contributions. However, it does not capture the utility of members' contributions. A more precise measure should include the interest of the members in participating with users from different clusters or communities, avoiding the rise of closed communities.

Contributory systems designers can take advantage of the results shown in Section 4.3 in order to design their regulation mechanisms, and adapt the effort evaluation of each member to the participatory layer robustness.

Acknowledgments

This work was partially supported by the European Community through the projects Community Networks Testbed for the Future Internet (CONFINE): FP7-288535, A Community Networking Cloud in a Box (CLOMMUNITY): FP7-317879, and also by the Spanish government under contract TIN2013-47245-C2-1-R.

References

1. FunkFeuer free net, <http://www.funkfeuer.at>
2. Guifi.net: Open, Free and Neutral Network Internet for everybody, <http://guifi.net/en>
3. Avonts, J., Braem, B., Blondia, C.: A questionnaire based examination of community networks. In: Wireless and Mobile Computing, Networking and Communications. pp. 8–15. (WiMob), IEEE (2013)
4. Bird, C., Gourley, A., Devanbu, P., Gertz, M., Swaminathan, A.: Mining email social networks. In: Mining Software Repositories. pp. 137–143. ACM (2006)
5. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008(10) (2008)
6. Braem, B., Baig Viñas, R., Kaplan, A.L., Neumann, A., et al.: A case for research with and on community networks. *ACM SIGCOMM Computer Communication Review* 43(3), 68–73 (2013)
7. Dorat, R., Latapy, M., Conein, B., Auray, N.: Multi-level analysis of an interaction network between individuals in a mailing-list. *Annales Des Telecommunications* 62(3-4), 325–349 (2007)
8. Gini, C.: On the Measure of Concentration with Especial Reference to Income and Wealth. Cowles Commission (1936)
9. Hristova, D., Musolesi, M., Mascolo, C.: Keep your friends close and your facebook friends closer: A multiplex network approach to the analysis of offline and online social ties. In: Weblogs and Social Media (2014)
10. Iriberry, A., Leroy, G.: A life-cycle perspective on online community success. *ACM Computing Surveys* 41(2), 11:1–11:29 (2009)
11. Kivelä, M., Arenas, A., Barthélemy, M., Gleeson, J.P., Moreno, Y., Porter, M.A.: Multilayer networks. *CoRR* abs/1309.7233 (2013)
12. Kleinberg, J.M.: Hubs, authorities, and communities. *ACM Comput. Surv.* 31(4es) (1999)
13. Marinos, A., Briscoe, G.: Community cloud computing. In: Jaatun, M.G., Zhao, G., Rong, C. (eds.) *Conference on Cloud Computing. LNCS*, vol. 5931, pp. 472–484. Springer (2009)
14. Palla, G., Derényi, I., Farkas, I., Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435(7043), 814–818 (2005)
15. Pfitzmann, A., Köhntopp, M.: Anonymity, unobservability, and pseudonymity - a proposal for terminology. In: Federrath, H. (ed.) *Workshop on Design Issues in Anonymity and Unobservability. LNCS*, vol. 2009, pp. 1–9. Springer (2001)
16. Vega, D., Meseguer, R., Ochoa, S., Pino, J., Freitag, F., et al.: Sharing hardware resources in heterogeneous computer-supported collaboration scenarios. *Integrated Computer-Aided Engineering* 20(1), 59–77 (2013)